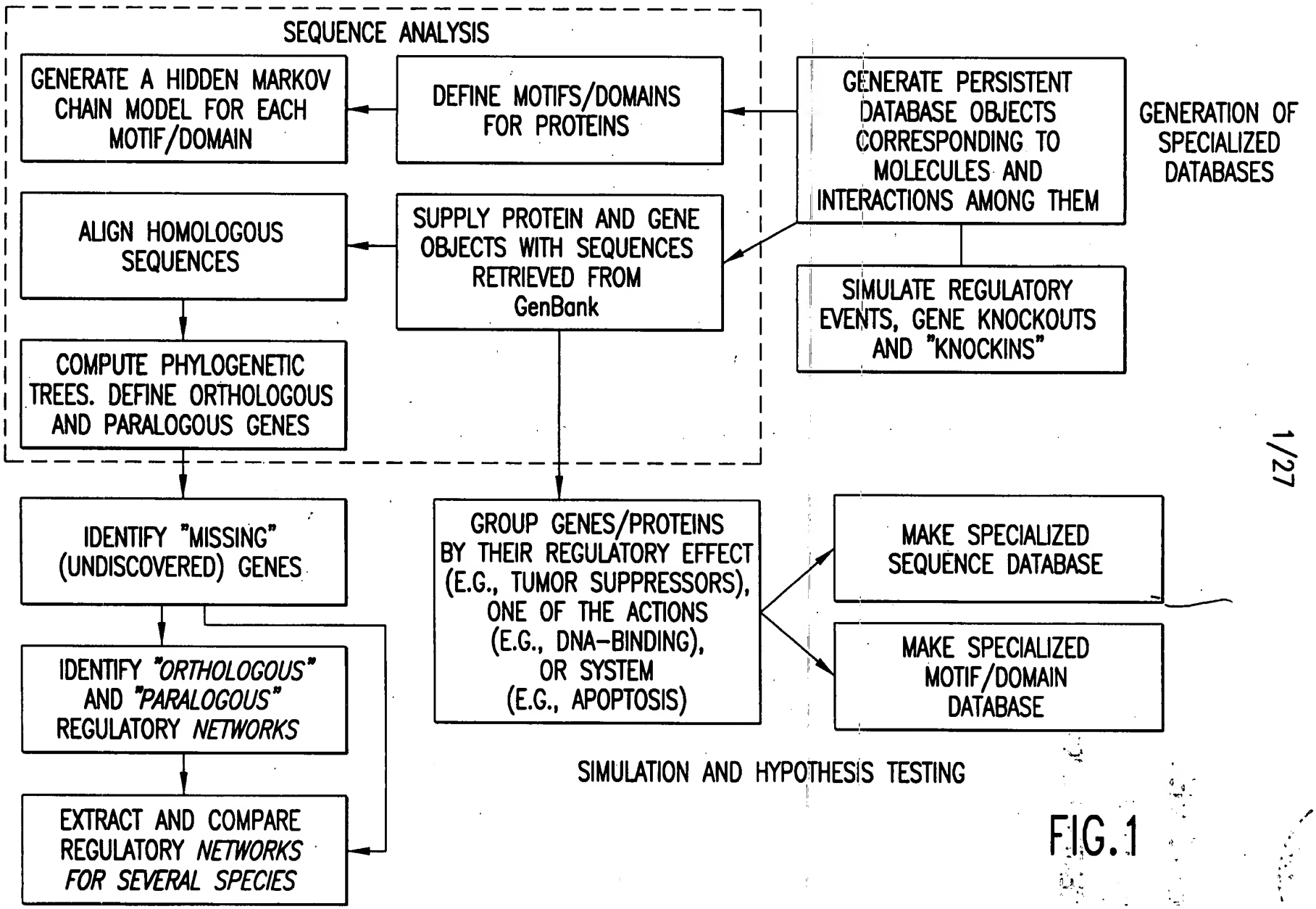

SEQUENCE ANALYSIS
GENERATE A HIDDEN MARKOV CHAIN MODEL FOR EACH MOTIF/DOMAIN
DEFINE MOTIFS/DOMAINS FOR PROTEINS
GENERATE PERSISTENT DATABASE OBJECTS CORRESPONDING TO MOLECULES AND INTERACTIONS AMONG THEM
GENERATION OF SPECIALIZED DATABASES
ALIGN HOMOLOGOUS SEQUENCES
SUPPLY PROTEIN AND GENE OBJECTS WITH SEQUENCES RETRIEVED FROM GenBank
SIMULATE REGULATORY EVENTS, GENE KNOCKOUTS AND "KNOCKINS"
COMPUTE PHYLOGENETIC TREES. DEFINE ORTHOLOGOUS AND PARALOGOUS GENES
IDENTIFY "MISSING" (UNDISCOVERED) GENES
GROUP GENES/PROTEINS BY THEIR REGULATORY EFFECT (E.G., TUMOR SUPPRESSORS), ONE OF THE ACTIONS (E.G., DNA-BINDING), OR SYSTEM (E.G., APOPTOSIS)
MAKE SPECIALIZED SEQUENCE DATABASE
MAKE SPECIALIZED MOTIF/DOMAIN DATABASE
IDENTIFY "ORTHOLOGOUS" AND "PARALOGOUS" REGULATORY NETWORKS
EXTRACT AND COMPARE REGULATORY NETWORKS FOR SEVERAL SPECIES
SIMULATION AND HYPOTHESIS TESTING

FIG.1

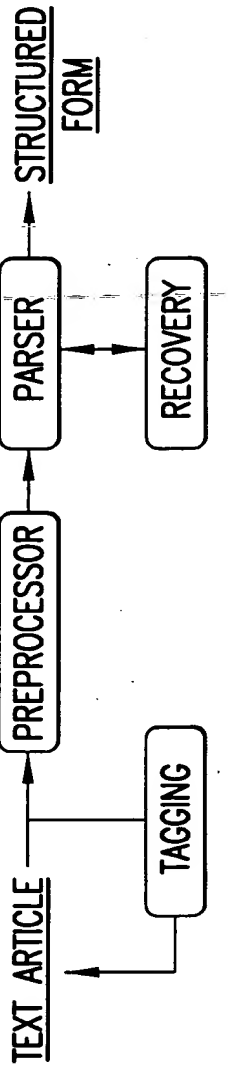

TEXT ARTICLE
TAGGING
LEXICON
PREPROCESSOR
GRAMMAR
PARSER
RECOVERY
STRUCTURED FORM

FIG.2

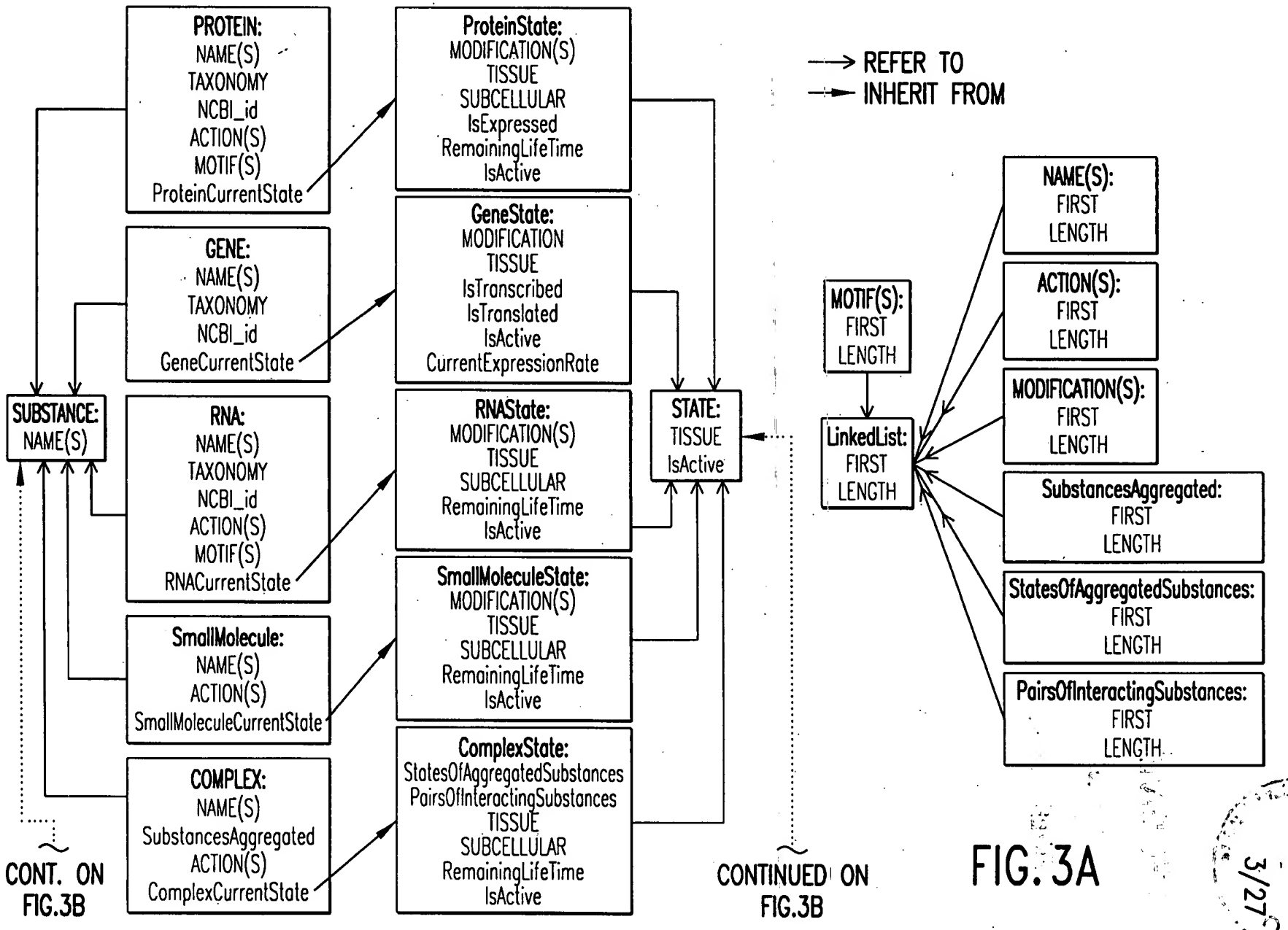

PROTEIN:
NAME(S)
TAXONOMY
NCBI_id
ACTION(S)
MOTIF(S)
ProteinCurrentState
GENE:
NAME(S)
TAXONOMY
NCBI_id
GeneCurrentState
RNA:
NAME(S)
TAXONOMY
NCBI_id
ACTION(S)
MOTIF(S)
RNACurrentState
SmallMolecule:
NAME(S)
ACTION(S)
SmallMoleculeCurrentState
COMPLEX:
NAME(S)
SubstancesAggregated
ACTION(S)
ComplexCurrentState
SUBSTANCE:
NAME(S)
ProteinState:
MODIFICATION(S)
TISSUE
SUBCELLULAR
IsExpressed
RemainingLifeTime
IsActive
GeneState:
MODIFICATION
TISSUE
IsTranscribed
IsTranslated
IsActive
CurrentExpressionRate
RNAState:
MODIFICATION(S)
TISSUE
SUBCELLULAR
RemainingLifeTime
IsActive
SmallMoleculeState:
MODIFICATION(S)
TISSUE
SUBCELLULAR
RemainingLifeTime
IsActive
ComplexState:
StatesOfAggregatedSubstances
PairsOfInteractingSubstances
TISSUE
SUBCELLULAR
RemainingLifeTime
IsActive
STATE:
TISSUE
IsActive
REFER TO
INHERIT FROM
MOTIF(S):
FIRST
LENGTH
LinkedList:
FIRST
LENGTH
NAME(S):
FIRST
LENGTH
ACTION(S):
FIRST
LENGTH
MODIFICATION(S):
FIRST
LENGTH
SubstancesAggregated:
FIRST
LENGTH
StatesOfAggregatedSubstances:
FIRST
LENGTH
PairsOfInteractingSubstances:
FIRST
LENGTH
CONT. ON
FIG.3B
CONTINUED ON
FIG.3B

FIG. 3A

3/27

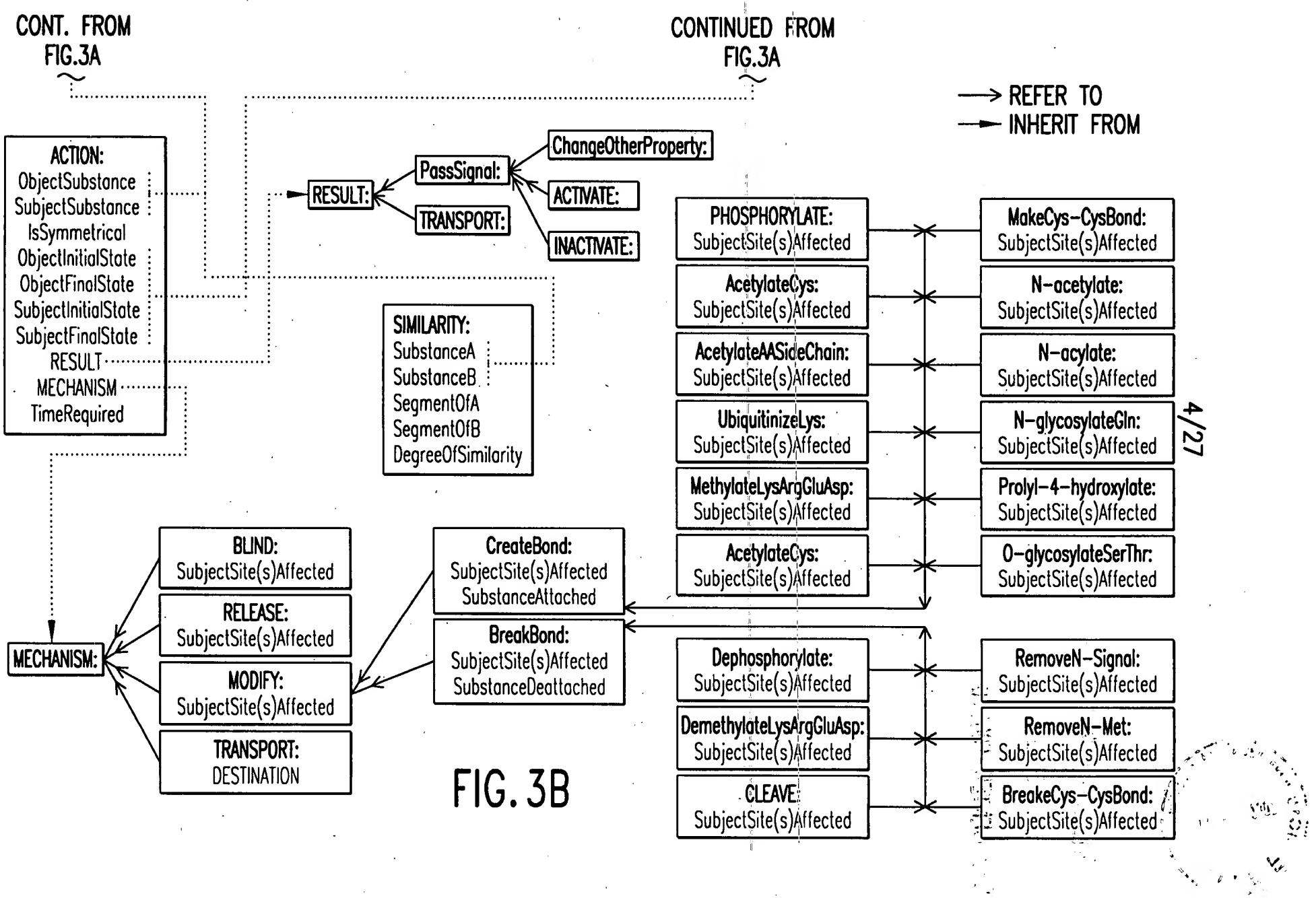

CONT. FROM FIG.3A
CONTINUED FROM FIG.3A
REFER TO
INHERIT FROM
ACTION:
ObjectSubstance
SubjectSubstance
IsSymmetrical
ObjectInitialState
ObjectFinalState
SubjectInitialState
SubjectFinalState
RESULT
MECHANISM
TimeRequired
RESULT:
PassSignal:
ChangeOtherProperty:
ACTIVATE:
TRANSPORT:
INACTIVATE:
SIMILARITY:
SubstanceA
SubstanceB
SegmentOfA
SegmentOfB
DegreeOfSimilarity
MECHANISM:
BLIND:
SubjectSite(s)Affected
RELEASE:
SubjectSite(s)Affected
MODIFY:
SubjectSite(s)Affected
TRANSPORT:
DESTINATION
CreateBond:
SubjectSite(s)Affected
SubstanceAttached
BreakBond:
SubjectSite(s)Affected
SubstanceDeattached
PHOSPHORYLATE:
SubjectSite(s)Affected
AcetylateCys:
SubjectSite(s)Affected
AcetylateAASideChain:
SubjectSite(s)Affected
UbiquitinizeLys:
SubjectSite(s)Affected
MethylateLysArgGluAsp:
SubjectSite(s)Affected
AcetylateCys:
SubjectSite(s)Affected
Dephosphorylate:
SubjectSite(s)Affected
DemethylateLysArgGluAsp:
SubjectSite(s)Affected
CLEAVE:
SubjectSite(s)Affected
MakeCys-CysBond:
SubjectSite(s)Affected
N-acetylate:
SubjectSite(s)Affected
N-acylate:
SubjectSite(s)Affected
N-glycosylateGln:
SubjectSite(s)Affected
Prolyl-4-hydroxylate:
SubjectSite(s)Affected
O-glycosylateSerThr:
SubjectSite(s)Affected
RemoveN-Signal:
SubjectSite(s)Affected
RemoveN-Met:
SubjectSite(s)Affected
BreakeCys-CysBond:
SubjectSite(s)Affected
4/27

FIG. 3B

bci-xL / bci / bci-xS / ced-9 / Bax / Blk / Bak / p21 / NGFI-B / N10 /Nak1 / Nur77 / Nurr 1 / Nor-1 / Noi-1 / RXR/ galectin-1 / N-glycan / CNTF / lck / fyn / ZAP-70 / raf / ras / MAP / protein kinase C / PKC / phosphatase calcineurin / NF-AT / AP1 / 14-3-3 / Raf-1 / Bci-2 / interieukin / IL-1 / IL-3 / cytokine / IGF-1 / CD95 / Apo-1 / RIP / FAF1 / FADD / FAP-1 / TNFR / TRAF / TRAP1 / TRAP2 / TRADD / HIAP1 / HIAP2 / CD40 / CD30 / XIAP / CD2 / CD3 / TCR / Bci-w / Mci-1 / NR-13 / BHRF1 / HMW5-HL / E1B19K / Nbk / Mch2 / CPP32 / ICE / FLICE / Nedd-2 / TX / Mch3 / Mch4 / ICB-1s / nor-1 / DNAseI / caspase / MACH1 / Mch5 / apopain / Yama / ICH / CMH / ced-3 / ced-4 / ced-9 / p53 / MKK3 / MKK1 / MKK2 / MKK4 / BAG-1 / Src / FAST/ p38 / p42 / ERK1 / p44 / ERK2 / SAPK / JNK / MEK / C-JUN / MEF2D / ATF2 / calcineurin / ELK-1 / protein phosphatase 2A / raf-1 / IL-I beta / TNF / PTK / Apaf / p35 / ETS / C-Myc / IL-2 / IL-2 receptor / NF-kappa B / TNFR-1 / TRAIL / APO-2L / DR4 / death receptor / DR3 / DR2 / DR5 / DR1 / bod / BMPR / BMP-x / TGF / grim / bid / FAN / perforin / Fas-L / Fas / DcR1 / decoy receptor / wxi-1 / NGF receptor / growth factor / RAR

FIG.4

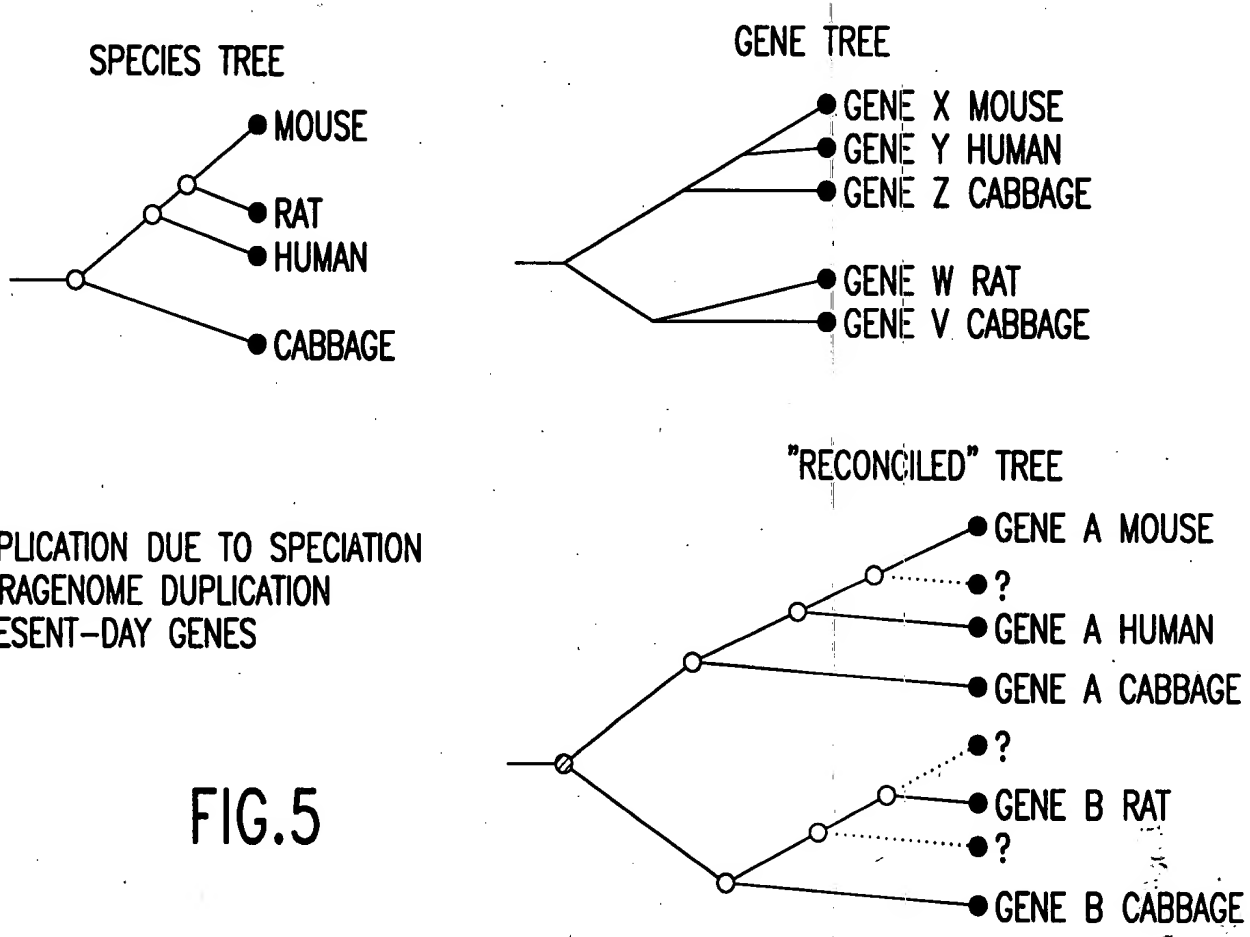

SPECIES TREE
MOUSE
RAT
HUMAN
CABBAGE
GENE TREE
GENE X MOUSE
GENE Y HUMAN
GENE Z CABBAGE
GENE W RAT
GENE V CABBAGE
"RECONCILED" TREE
GENE A MOUSE
?
GENE A HUMAN
GENE A CABBAGE
?
GENE B RAT
?
GENE B CABBAGE
○DUPLICATION DUE TO SPECIATION
⊘INTRAGENOME DUPLICATION
PRESENT-DAY GENES

FIG.5

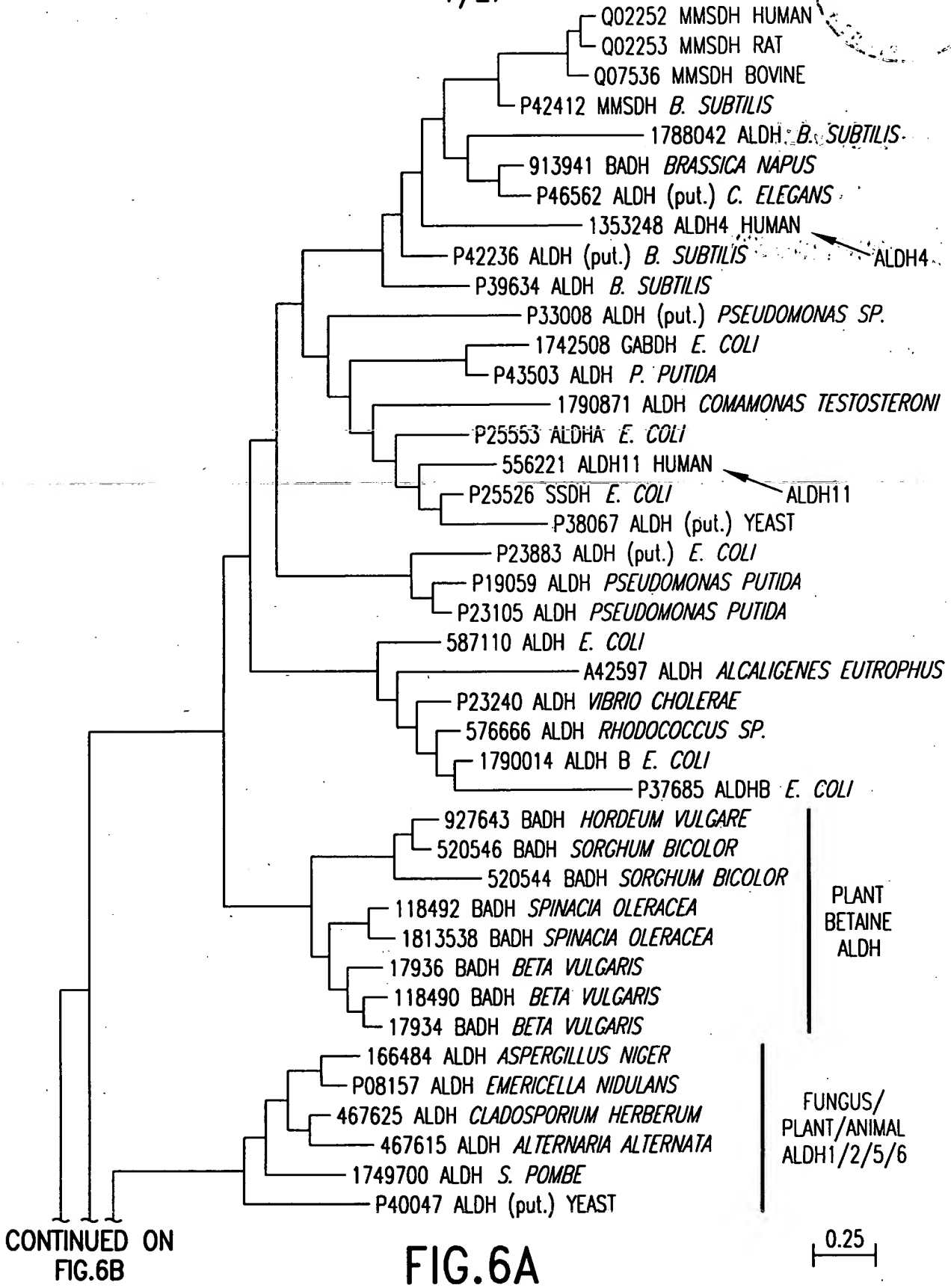

Q02252 MMSDH HUMAN
Q02253 MMSDH RAT
Q07536 MMSDH BOVINE
P42412 MMSDH B. SUBTILIS
1788042 ALDH B. SUBTILIS
913941 BADH BRASSICA NAPUS
P46562 ALDH (put.) C. ELEGANS
1353248 ALDH4 HUMAN
ALDH4
P42236 ALDH (put.) B. SUBTILIS
P39634 ALDH B. SUBTILIS
P33008 ALDH (put.) PSEUDOMONAS SP.
1742508 GABDH E. COLI
P43503 ALDH P. PUTIDA
1790871 ALDH COMAMONAS TESTOSTERONI
P25553 ALDHA E. COLI
556221 ALDH11 HUMAN
ALDH11
P25526 SSDH E. COLI
P38067 ALDH (put.) YEAST
P23883 ALDH (put.) E. COLI
P19059 ALDH PSEUDOMONAS PUTIDA
P23105 ALDH PSEUDOMONAS PUTIDA
587110 ALDH E. COLI
A42597 ALDH ALCALIGENES EUTROPHUS
P23240 ALDH VIBRIO CHOLERAE
576666 ALDH RHODOCOCCUS SP.
1790014 ALDH B E. COLI
P37685 ALDHB E. COLI
927643 BADH HORDEUM VULGARE
520546 BADH SORGHUM BICOLOR
520544 BADH SORGHUM BICOLOR
118492 BADH SPINACIA OLERACEA
1813538 BADH SPINACIA OLERACEA
17936 BADH BETA VULGARIS
118490 BADH BETA VULGARIS
17934 BADH BETA VULGARIS
PLANT BETAINE ALDH
166484 ALDH ASPERGILLUS NIGER
P08157 ALDH EMERICELLA NIDULANS
467625 ALDH CLADOSPORIUM HERBERUM
467615 ALDH ALTERNARIA ALTERNATA
1749700 ALDH S. POMBE
P40047 ALDH (put.) YEAST
FUNGUS/ PLANT/ANIMAL ALDH1/2/5/6
CONTINUED ON FIG.6B
0.25

FIG.6A

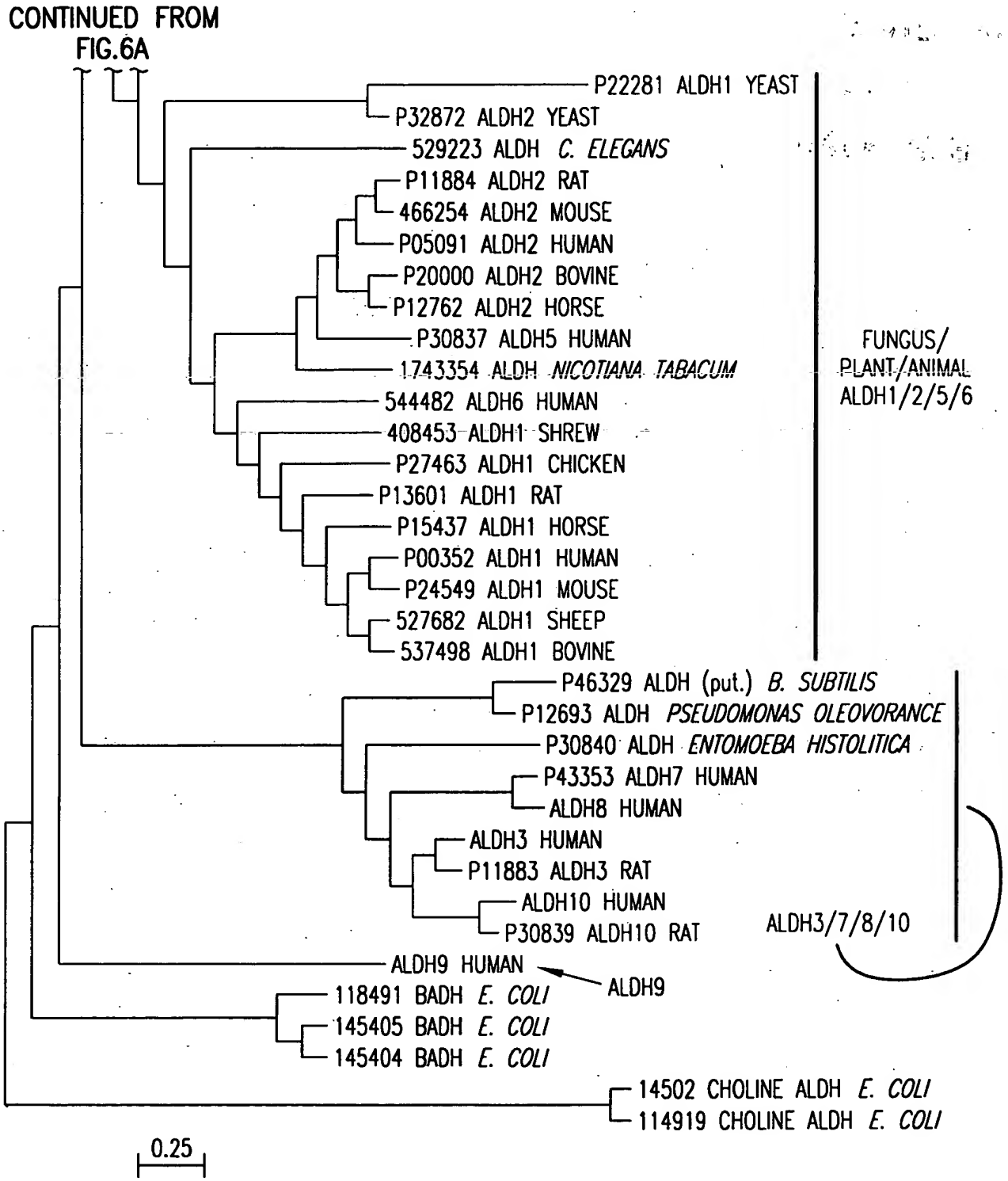

CONTINUED FROM FIG.6A
P22281 ALDH1 YEAST
P32872 ALDH2 YEAST
529223 ALDH C. ELEGANS
P11884 ALDH2 RAT
466254 ALDH2 MOUSE
P05091 ALDH2 HUMAN
P20000 ALDH2 BOVINE
P12762 ALDH2 HORSE
P30837 ALDH5 HUMAN
1743354 ALDH NICOTIANA TABACUM
544482 ALDH6 HUMAN
408453 ALDH1 SHREW
P27463 ALDH1 CHICKEN
P13601 ALDH1 RAT
P15437 ALDH1 HORSE
P00352 ALDH1 HUMAN
P24549 ALDH1 MOUSE
527682 ALDH1 SHEEP
537498 ALDH1 BOVINE
FUNGUS/ PLANT/ANIMAL ALDH1/2/5/6
P46329 ALDH (put.) B. SUBTILIS
P12693 ALDH PSEUDOMONAS OLEOVORANCE
P30840 ALDH ENTOMOEBA HISTOLITICA
P43353 ALDH7 HUMAN
ALDH8 HUMAN
ALDH3 HUMAN
P11883 ALDH3 RAT
ALDH10 HUMAN
P30839 ALDH10 RAT
ALDH3/7/8/10
ALDH9 HUMAN
ALDH9
118491 BADH E. COLI
145405 BADH E. COLI
145404 BADH E. COLI
14502 CHOLINE ALDH E. COLI
114919 CHOLINE ALDH E. COLI
0.25

FIG.6B

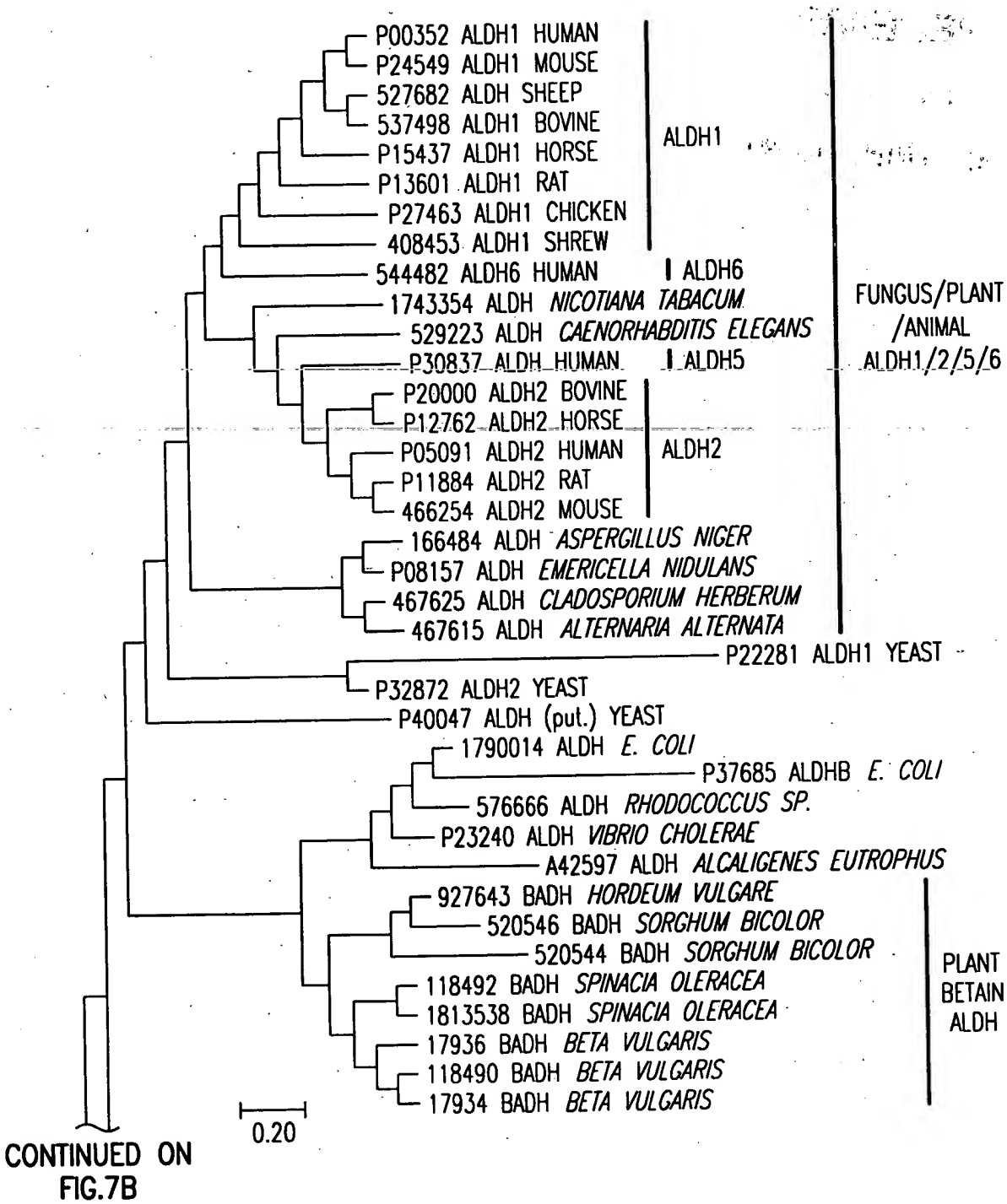

P00352 ALDH1 HUMAN
P24549 ALDH1 MOUSE
527682 ALDH SHEEP
537498 ALDH1 BOVINE
P15437 ALDH1 HORSE
P13601 ALDH1 RAT
P27463 ALDH1 CHICKEN
408453 ALDH1 SHREW
ALDH1
544482 ALDH6 HUMAN
ALDH6
1743354 ALDH NICOTIANA TABACUM
529223 ALDH CAENORHABDITIS ELEGANS
P30837 ALDH HUMAN
ALDH5
FUNGUS/PLANT /ANIMAL ALDH1/2/5/6
P20000 ALDH2 BOVINE
P12762 ALDH2 HORSE
P05091 ALDH2 HUMAN
P11884 ALDH2 RAT
466254 ALDH2 MOUSE
ALDH2
166484 ALDH ASPERGILLUS NIGER
P08157 ALDH EMERICELLA NIDULANS
467625 ALDH CLADOSPORIUM HERBERUM
467615 ALDH ALTERNARIA ALTERNATA
P22281 ALDH1 YEAST
P32872 ALDH2 YEAST
P40047 ALDH (put.) YEAST
1790014 ALDH E. COLI
P37685 ALDHB E. COLI
576666 ALDH RHODOCOCCUS SP.
P23240 ALDH VIBRIO CHOLERAE
A42597 ALDH ALCALIGENES EUTROPHUS
927643 BADH HORDEUM VULGARE
520546 BADH SORGHUM BICOLOR
520544 BADH SORGHUM BICOLOR
118492 BADH SPINACIA OLERACEA
1813538 BADH SPINACIA OLERACEA
17936 BADH BETA VULGARIS
118490 BADH BETA VULGARIS
17934 BADH BETA VULGARIS
PLANT BETAIN ALDH
0.20
CONTINUED ON FIG.7B

FIG. 7A

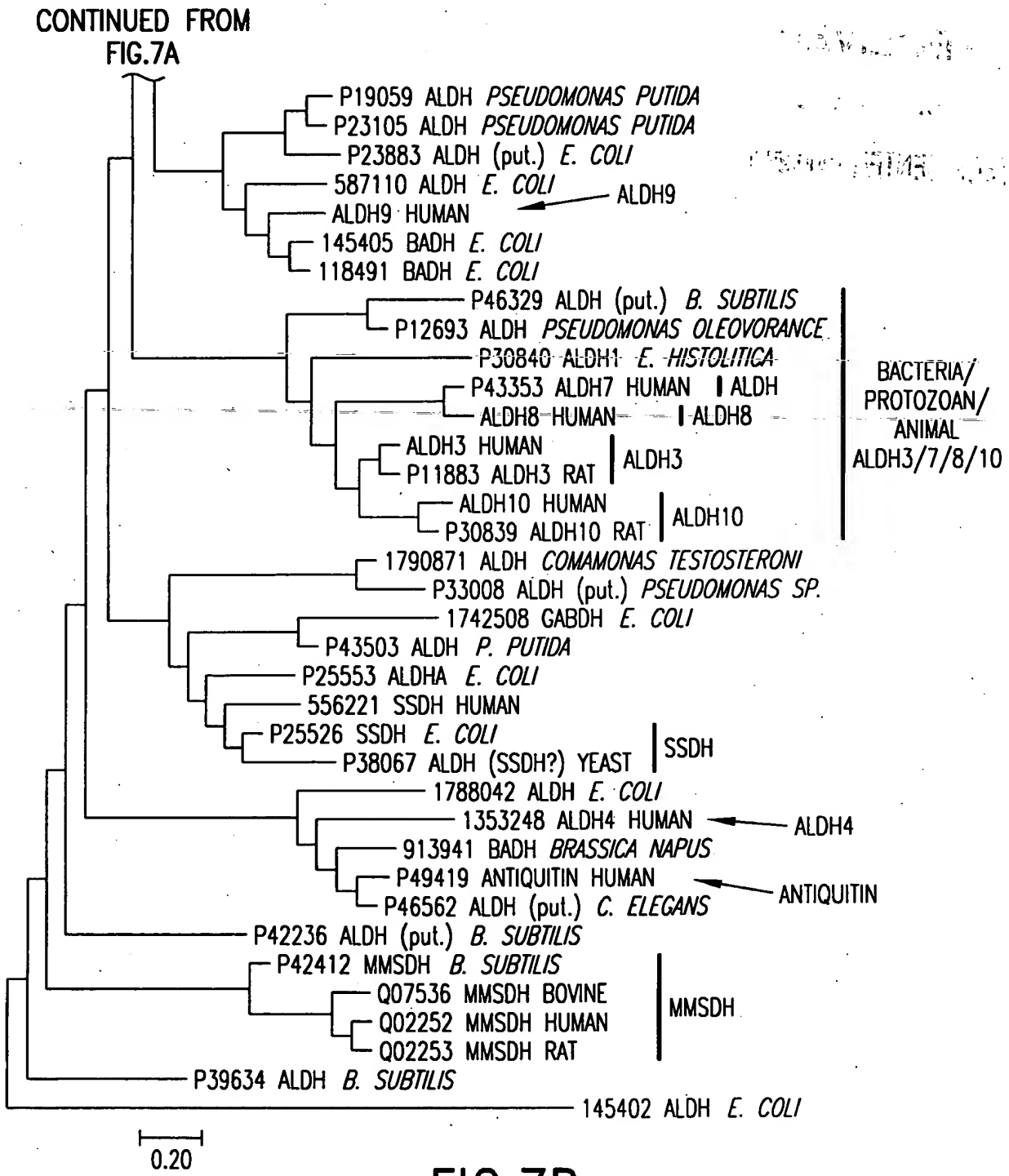

CONTINUED FROM FIG.7A
P19059 ALDH PSEUDOMONAS PUTIDA
P23105 ALDH PSEUDOMONAS PUTIDA
P23883 ALDH (put.) E. COLI
587110 ALDH E. COLI
ALDH9 HUMAN
ALDH9
145405 BADH E. COLI
118491 BADH E. COLI
P46329 ALDH (put.) B. SUBTILIS
P12693 ALDH PSEUDOMONAS OLEOVORANCE
P30840 ALDH1 E. HISTOLITICA
P43353 ALDH7 HUMAN
ALDH
ALDH8 HUMAN
ALDH8
ALDH3 HUMAN
P11883 ALDH3 RAT
ALDH3
ALDH10 HUMAN
P30839 ALDH10 RAT
ALDH10
BACTERIA/ PROTOZOAN/ ANIMAL ALDH3/7/8/10
1790871 ALDH COMAMONAS TESTOSTERONI
P33008 ALDH (put.) PSEUDOMONAS SP.
1742508 GABDH E. COLI
P43503 ALDH P. PUTIDA
P25553 ALDHA E. COLI
556221 SSDH HUMAN
P25526 SSDH E. COLI
P38067 ALDH (SSDH?) YEAST
SSDH
1788042 ALDH E. COLI
1353248 ALDH4 HUMAN
ALDH4
913941 BADH BRASSICA NAPUS
P49419 ANTIQUITIN HUMAN
P46562 ALDH (put.) C. ELEGANS
ANTIQUITIN
P42236 ALDH (put.) B. SUBTILIS
P42412 MMSDH B. SUBTILIS
Q07536 MMSDH BOVINE
Q02252 MMSDH HUMAN
Q02253 MMSDH RAT
MMSDH
P39634 ALDH B. SUBTILIS
145402 ALDH E. COLI
0.20

FIG.7B

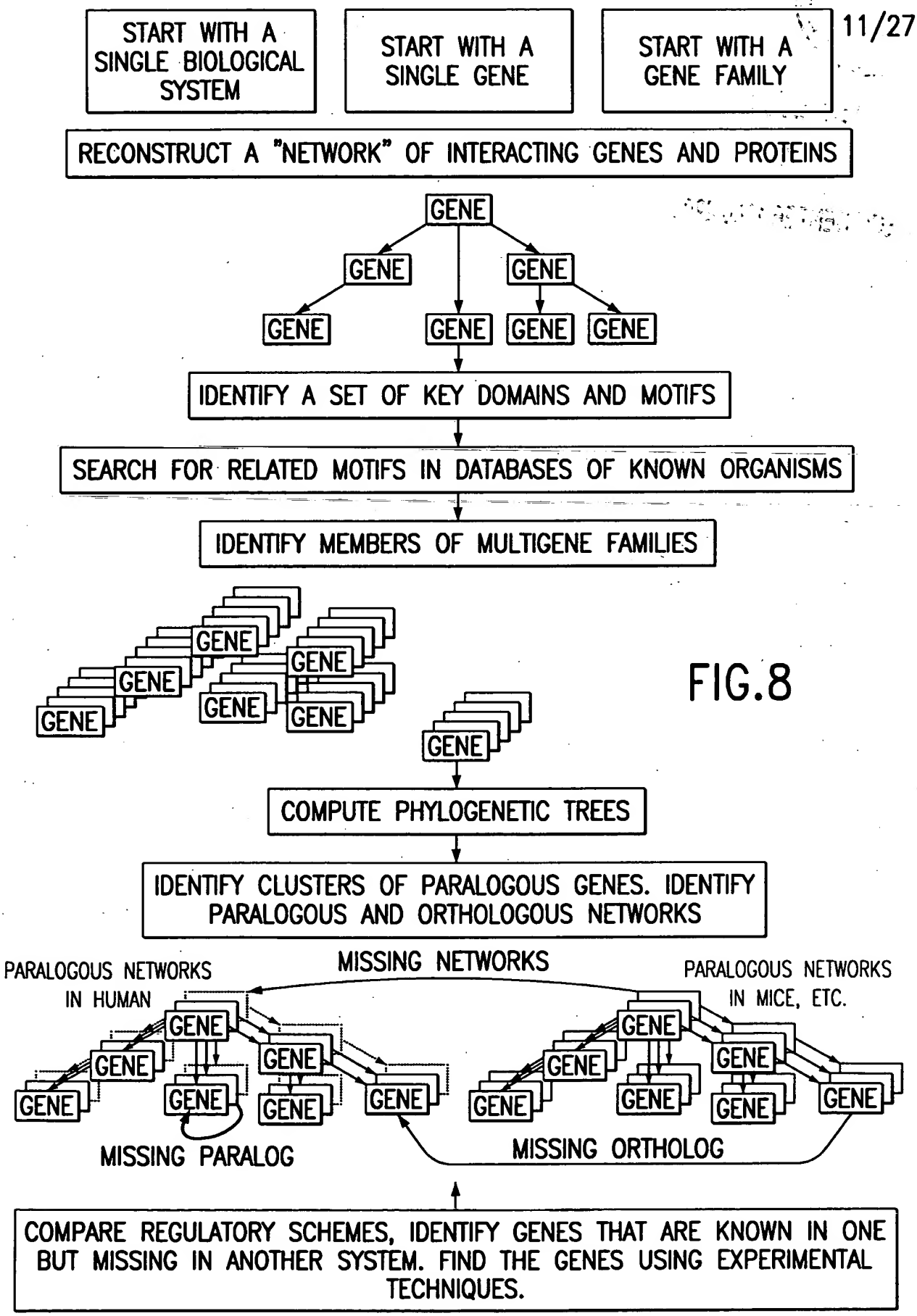

START WITH A SINGLE BIOLOGICAL SYSTEM
START WITH A SINGLE GENE
START WITH A GENE FAMILY
RECONSTRUCT A "NETWORK" OF INTERACTING GENES AND PROTEINS
GENE
IDENTIFY A SET OF KEY DOMAINS AND MOTIFS
SEARCH FOR RELATED MOTIFS IN DATABASES OF KNOWN ORGANISMS
IDENTIFY MEMBERS OF MULTIGENE FAMILIES
COMPUTE PHYLOGENETIC TREES
IDENTIFY CLUSTERS OF PARALOGOUS GENES. IDENTIFY PARALOGOUS AND ORTHOLOGOUS NETWORKS
PARALOGOUS NETWORKS IN HUMAN
MISSING NETWORKS
PARALOGOUS NETWORKS IN MICE, ETC.
MISSING PARALOG
MISSING ORTHOLOG
COMPARE REGULATORY SCHEMES, IDENTIFY GENES THAT ARE KNOWN IN ONE BUT MISSING IN ANOTHER SYSTEM. FIND THE GENES USING EXPERIMENTAL TECHNIQUES.

FIG.8

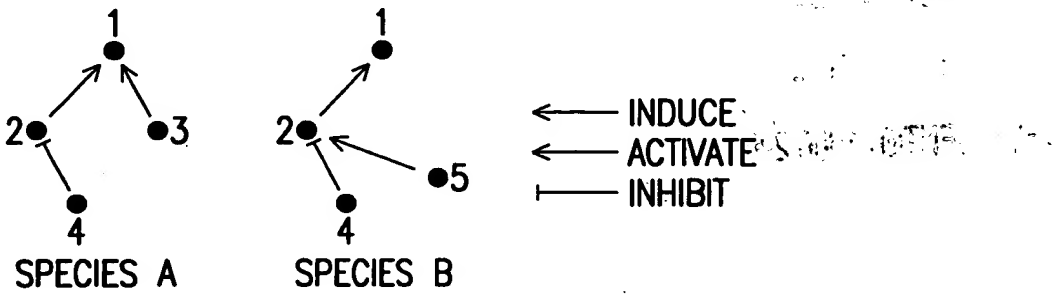


FIG.9A

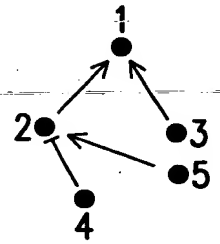


FIG.9B

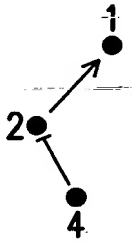


FIG.9C

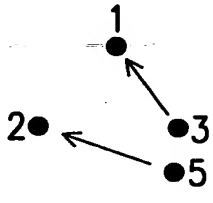


FIG.9D

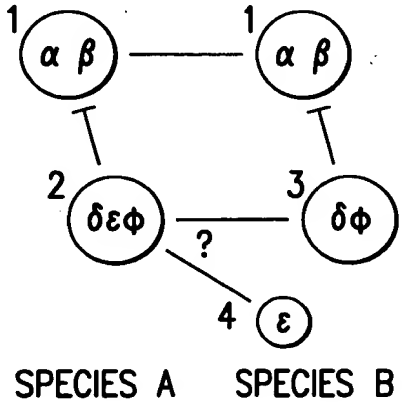


FIG.10

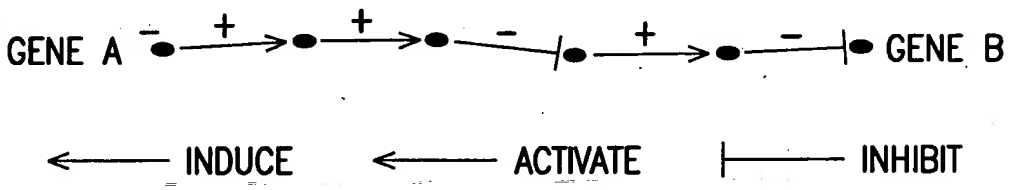


FIG.11A

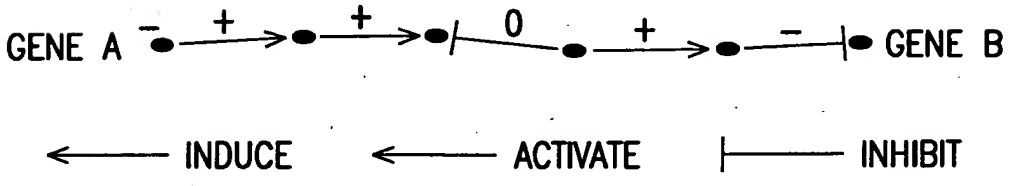


FIG.11B

FLOW CHART SCHEME OF GENE DISCOVERY ANALYSIS INVOLVING MOTIF/DOMAIN ANALYSIS.

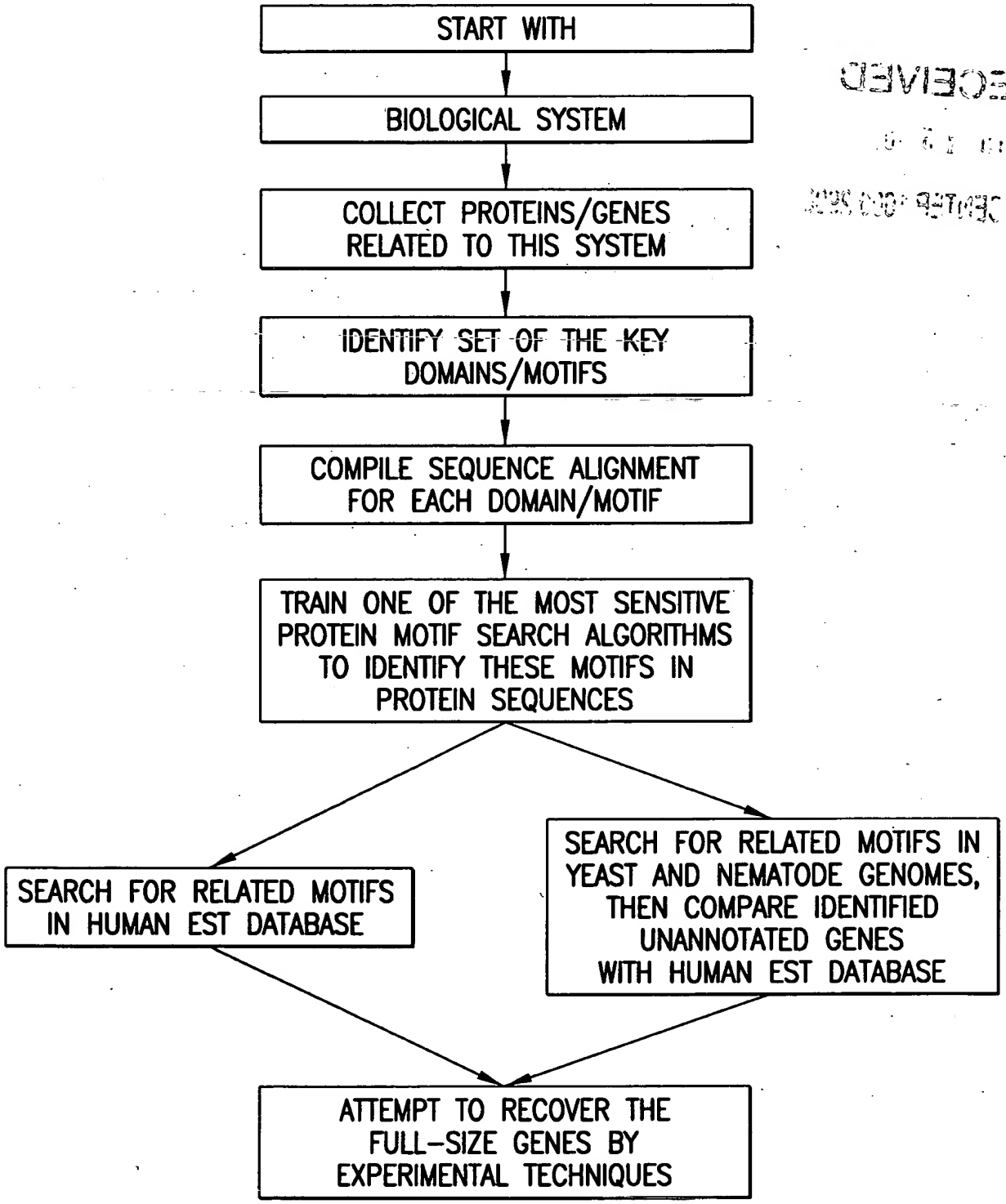


FIG.12

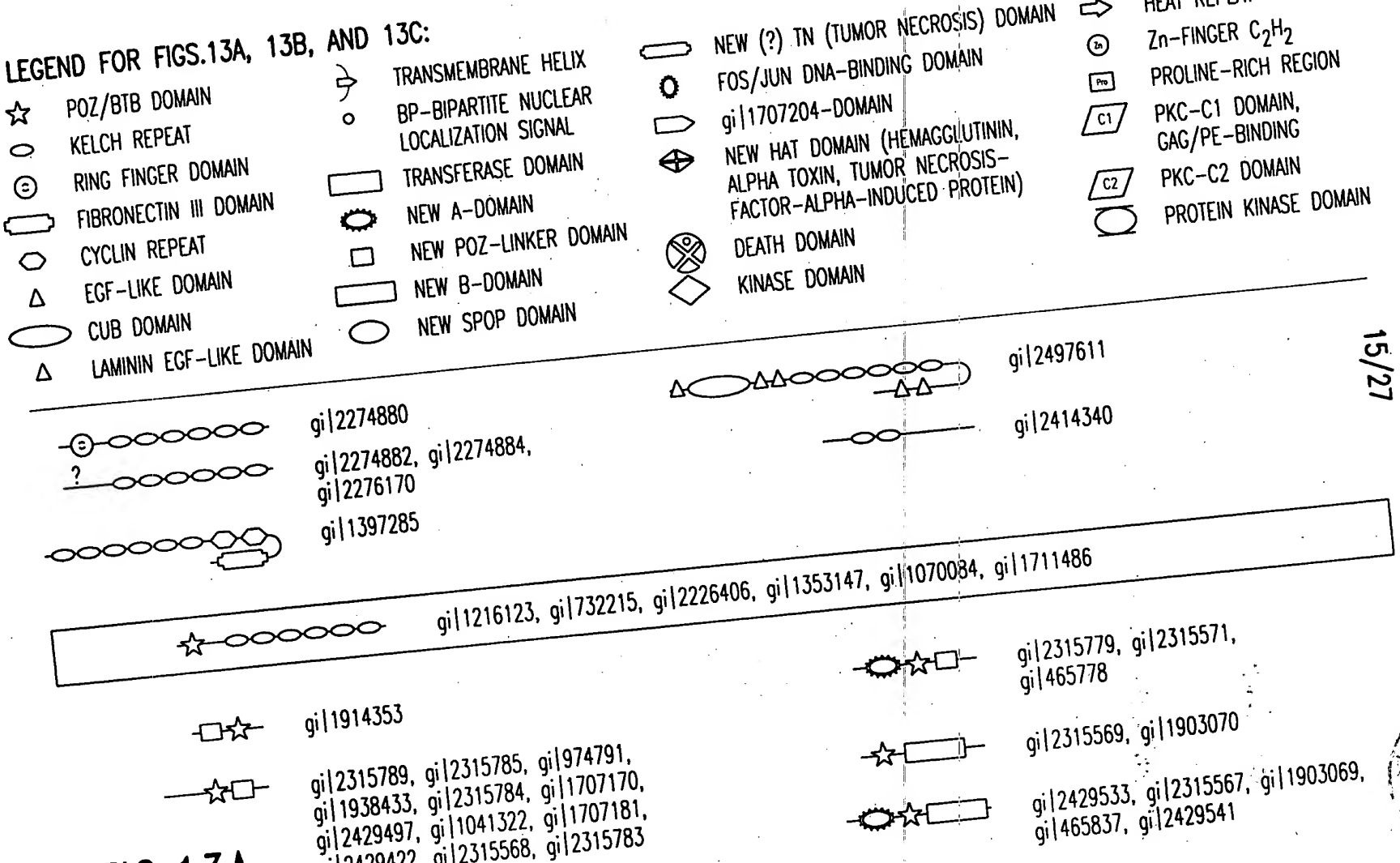
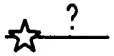
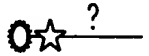

LEGEND FOR FIGS.13A, 13B, AND 13C:
POZ/BTB DOMAIN
KELCH REPEAT
RING FINGER DOMAIN
FIBRONECTIN III DOMAIN
CYCLIN REPEAT
EGF-LIKE DOMAIN
CUB DOMAIN
LAMININ EGF-LIKE DOMAIN
TRANSMEMBRANE HELIX
BP-BIPARTITE NUCLEAR LOCALIZATION SIGNAL
TRANSFERASE DOMAIN
NEW A-DOMAIN
NEW POZ-LINKER DOMAIN
NEW B-DOMAIN
NEW SPOP DOMAIN
NEW (?) TN (TUMOR NECROSIS) DOMAIN
FOS/JUN DNA-BINDING DOMAIN
gi|1707204-DOMAIN
NEW HAT DOMAIN (HEMAGGLUTININ, ALPHA TOXIN, TUMOR NECROSIS-FACTOR-ALPHA-INDUCED PROTEIN)
DEATH DOMAIN
KINASE DOMAIN
HEAT REPEAT
Zn-FINGER C_2H_2
PROLINE-RICH REGION
PKC-C1 DOMAIN, GAG/PE-BINDING
PKC-C2 DOMAIN
PROTEIN KINASE DOMAIN
gi|2274880
gi|2274882, gi|2274884, gi|2276170
gi|1397285
gi|2497611
gi|2414340
gi|1216123, gi|732215, gi|2226406, gi|1353147, gi|1070084, gi|1711486
gi|1914353
gi|2315789, gi|2315785, gi|974791, gi|1938433, gi|2315784, gi|1707170, gi|2429497, gi|1041322, gi|1707181, gi|2429422, gi|2315568, gi|2315783
gi|2315779, gi|2315571, gi|465778
gi|2315569, gi|1903070
gi|2429533, gi|2315567, gi|1903069, gi|465837, gi|2429541

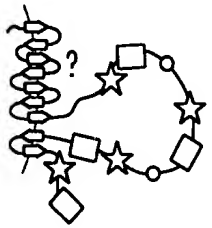
FIG.13A



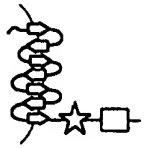
gi|1465836, gi|1707203, gi|2315752, gi|465779, gi|1707205, gi|2429493, gi|1707215, gi|1070062, gi|2315655



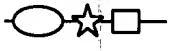
gi|2291257



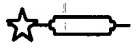
gi|2394485



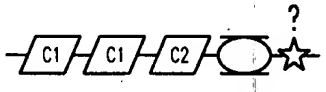
gi|1132514



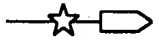
gi|529718, gi|1526968 (MEL26), gi|466032, gi|1176717



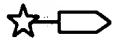
gi|1707217, gi|1707213, gi|1707214, gi|2315750, gi|2315635, gi|1707216, gi|2315748, gi|1707212, gi|1707202, gi|2315636, gi|2315655, gi|2315634, gi|2315541, gi|2315634



gi|868172



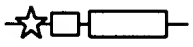
gi|2315751



gi|1707204, gi|1707206, gi 2315660, gi|2315661, gi|2315757 (POZ IS TRUNCATED)

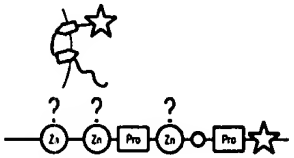
gi|522131

FIG.13B

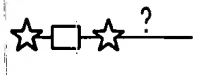


gi|2497016

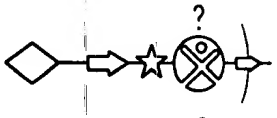
gi|1914354, gi|1293841, gi|1326360,
gi|1326361 (B0496.2), gi|1065930,
gi|2315541, gi|1049377, gi|1049378,
gi|1132541, gi|1903102, gi|1122793,
gi|1086886, gi|1729689, gi|1256456,
gi|1438721, gi|1256451, gi|1429312



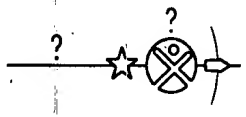
gi|2315652



gi|2315688

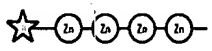


gi|1132541



gi|1469056

gi|2315570 (POZ IS TRUNCATED),
gi|2315788, gi|2429424



gi|671825

FIG.13C

>gi|2210766|gb|AA481214|AA481214 aa34e02.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:815162 5' similar to WP:W07A12.4 CE03795 :, mRNA sequence [Homo sapiens]
CATGGCTTCCTGGACACCAACCCTGCCATCCGGGAGCAGACGGTCAAGTCCATGCTGCTCCTGGCCCCAA
AGCTGAACGAGGCCAACCTCAATGTGGAGCTGATGAAGCACTTTGCACGGCTACAGGCCAAGGATGAACA
GGGCCCCATCCGCTGCAACACCACAGTCTGCCTGGGCAAAATCGGCTCCTACCTCAGTGCTAGCACCAGA
CACAGGGTCCTTACCTCTGCCTTCAGCCGAGCCACTAGGGACCCGTTTGCACCGTCCCGGGTTGCGGGTG
TCCTGGGCTTTGCTGCCACCCACAACCTCTACTCAATGAACGACTGTGCCCAGAAGATCCTGCCTGTGCT
CTGCGGTCTCACTGTAGATCCTGAGAAATCCGTGCGAGACCAGGCCTTCAAGGCA

>gi|1349211|gb|W51957|W51957 zc45f01.r1 Soares_senescent_fibroblasts_NbHSF Homo sapiens cDNA clone IMAGE:325273 5', mRNA sequence [Homo sapiens]
CCTTCGAGTTCGGCAATGCTGGGGCCGTTGTCCTCACGCCCCTCTTCAAGGTGGGCAAGTTCCTGAGCGC
TGAGGAGTATCAGCAGAAGATCATCCCTGTGGTGGTCAAGATGTTCTCATCCACTGACCGGGCCATGCGC
ATCCGNCTCCTGCAGCAGATGGAGCAGTTCATCCAGTACCTTGACGAGCCAACAGTCAACACCCAGATCT
TCCCCCACGTCGTACATGGCTTCCTGGACACCAACCCTGCCATCCGGGAGCAGACGGTCAAGTCCATGCT
GCTCCTGGCCCCAAAGCTGAACGAGGCCAACCTCAATGTGGAGCTGATGAAGCACTTTGCACGGCTACAG
GCCAAGGATGAACAGGGCCCCATCCGCTGCAACACCACAGTCTGCCTGGGCAAAATCGGCTCCTACCTCA
GTGCTAGCACCAGACACAGGGTCCTTACCTCTG

FIG.14A

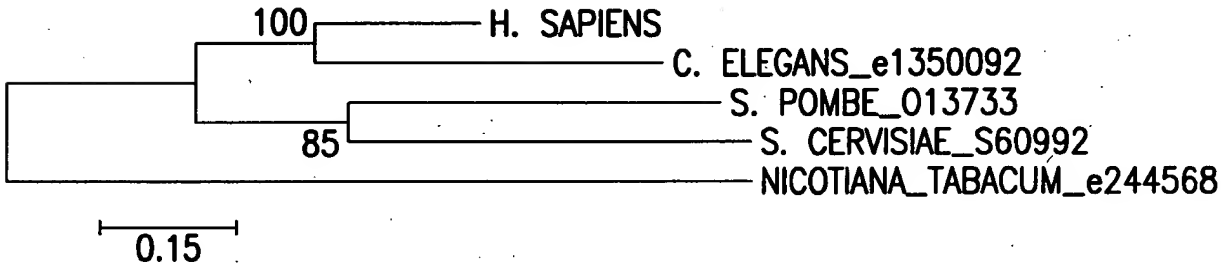

100
H. SAPIENS
C. ELEGANS_e1350092
S. POMBE_013733
85
S. CERVISIAE_S60992
NICOTIANA_TABACUM_e244568
0.15

FIG. 14B

BASE COUNT 405 a 545 c 493 g 278 t 6 others
ORIGIN

```
   1 cagccgaagc amgcaaaaat tcttccagga gctgagcaag agcctggacg cattccctga
  61 ggayttctgt cggcacaagg tgctgcccca gctgctgacc gccttcgagt tcggcaatgc
 121 tggggccgtt gtcctcacgc ccctcttcaa ggtgggcaag ttcctgagcg ctgaggagta
 181 tcagcagaag atcatccctg tggtggtcaa gatgttctca tccactgacc gggccatgcg
 241 catccgcctc ctgcagcaga tggagcagtt catccagtac cttgacgagc caacagtcaa
 301 cacccagatc ttcccccacg tcgtacatgg cttcctggac accaaccctg ccatccggga
 361 gcagacggtc aagtccatgc tgctcctggc cccaaagctg aacgaggcca acctcaatgt
 421 ggagctgatg aagcactttg cacggctaca ggccaaggat gaacagggcc ccatccgctg
 481 caacaccaca gtctgcctgg gcaaaatcgg ctcctacctc agtgctagca ccagacacag
 541 ggtccttacc tctgccttca gccgagccac tagggacccg tttgcaccgt cccgggttgc
 601 gggtgtcctg ggctttgctg ccacccacaa cctctactca atgaacgact gtgcccagaa
 661 gatcctgcct gtgctctgcg gtctcactgt agatcctgag aaatccgtgc gagaccaggc
 721 cttcaaggcm wttcggagct tcctgtccaa attggagtct gtgtcggagg acccgaccca
 781 gctggaggaa gtggagaagg atgtccatgc agcctccagc cctggcatgg gaggagccgc
 841 agctagctgg gcaggctggg cgtgaccggg gtctcctcac tcacctccaa gctgatccgt
 901 tcgcacccaa ccactgcccc aacagaaacc aacattcccc aaagacccac gcctgaagga
 961 gttcctgccc cagcccccac ccctgttcct gccaccccta caacctcagg ccactgggag
1021 acgcaggagg aggacaagga cacagcagag gacagcagca ctgctgacag atgggacgac
1081 gaagactggg gcagcctgga gcaggaggcc gagtctgtgc tggcccagca ggacgactgg
1141 agcaccgggg gccaagtgag ccgtgctagt caggtcagca actccgacca caaatcctcc
1201 aaatccccag agtccgactg gagcagctgg gaarctgagg gctcctggga acagggctgg
1261 caggagccaa gctcccagga gccacctyct gacggtacac ggctggccag cgagtataac
1321 tggggtggcc cagagtccag cgacaagggc gaccccttcg ctaccctgtc tgcacgtccc
1381 agcacccagc cgaggccaga ctcttggggt gaggacaact gggagggcct cgagactgac
1441 agtcgacagg tcaaggctga gctggcccgg aagaagcgcg aggagcggcg gcgggagatg
1501 gaggccaaac gcgccgagag gaaggtgcca agggccccat gaagctggga gcccggaagc
1561 tggactgaac cgtggcggtg gcccttcccg gctgcggaga gcccgcccca cagatgtatt
1621 tattgtacaa accatgtgag cccggccgcc cagccaggcc atctcacgtg tacataatca
1681 gagccacaat aaattctatt tcacaaaaaa aaaaaaaaaa aaaaaaa
```

FIG.14C

```
          5         10        15        20        25        30
  1  S R S X Q K F F Q E L S K S L D A F P E D F C R H K V L P Q
 31  L L T A F E F G N A G A V V L T P L F K V G K F L S A E E Y
 61  Q Q K I I P V V V K M F S S T D R A M R I R L L Q Q M E Q F
 91  I Q Y L D E P T V N T Q I F P H V V H G F L D T N P A I R E
121  Q T V K S M L L L A P K L N E A N L N V E L M K H F A R L Q
151  A K D E Q G P I R C N T T V C L G K I G S Y L S A S T R H R
181  V L T S A F S R A T R D P F A P S R V A G V L G F A A T H N
211  L Y S M N D C A Q K I L P V L C G L T V D P E K S V R D Q A
241  F K A X R S F L S K L E S V S E D P T Q L E E V E K D V H A
271  A S S P G M G G A A A S W A G W A
```

FIG.14D

RECEIVED

>sp|P15533|RPT1_MOUSE DOWN REGULATORY PROTEIN OF INTERLEUKIN 2 RECEPTOR (J03776) rpt-1r [Mus musculus] Length = 353

Score = 92.0 bits (237), Expect = 6e-20

```
Query 194  VMELLEEDLTCPICCSLFDDPRVLPCSHNFCKKCLEGILEGSVRNSMWRPAPFKCPTCRK  373
           V+E+++E++TCPIC  L  +P    C+H+FC+ C+    E S RN+        CP CR
Sbjct 5    VLEMIKEEVTCPICLELLKEPVSADCNHSFCRACITLNYE-SNRNT---DGKGNCPVCRV  60

Query 374  ETSATGINSLQVNYSLKGIVEKYNKIKISP----KMPVCKGHMGQPLNIFCLTDMQLICG  541
                  +L+ N  +  IVE+    K  P    K+ +C  H G+ L +FC   DM +IC
Sbjct 61   PYP---FGNLRPNLHVANIVERLKGFKSIPEEEQKVNICAQH-GEKLRLFCRKDMMVICW  116

Query 542  ICATRGEHTKHVFCSIEDAYAQERDAFESLFQSF-------ETWRRGDALSRLDTMETSK  700
           +C    EH  H    IE+   + ++  +            + W+    L R+D
Sbjct 117  LCERSQEHRGHQTALIEEVDQEYKEKLQGALWKLMKKAKICDEWQDDLQLQRVDW-----  171

Query 701  RKSLQLMTKDSDKVKEFFEKLQHTLDQKKNEILSDFETMKLAVMQAYDPEINKL  862
              +Q+   + + V+  F+ L+  LD K+NE L   +  K  VM+  +   N+L
Sbjct 172  ENQIQI---NVENVQRQFKGLRDLLDSKENEELQKLKKEKKEVMEKLEESENEL  222
```

FIG.15

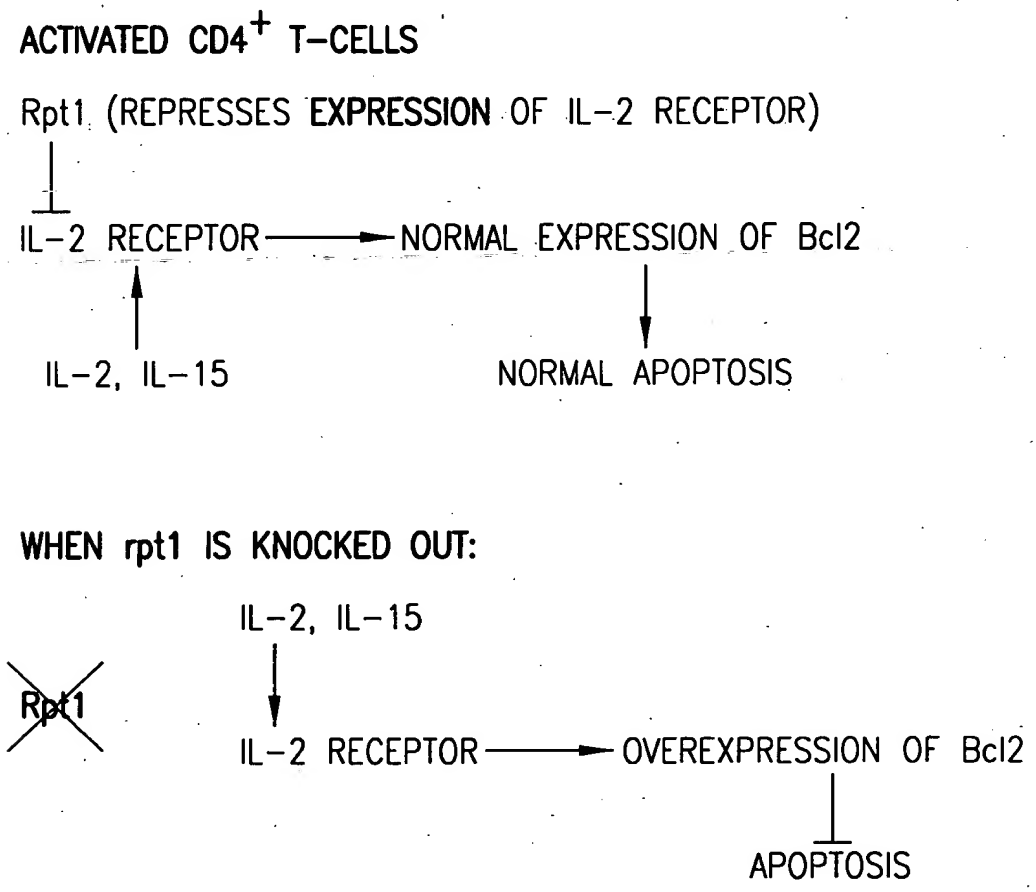

ACTIVATED CD4+ T-CELLS
Rpt1 (REPRESSES EXPRESSION OF IL-2 RECEPTOR)
IL-2 RECEPTOR
NORMAL EXPRESSION OF Bcl2
IL-2, IL-15
NORMAL APOPTOSIS
WHEN rpt1 IS KNOCKED OUT:
IL-2, IL-15
Rpt1
IL-2 RECEPTOR
OVEREXPRESSION OF Bcl2
APOPTOSIS

FIG. 16

```
Query= gi |2137498|Mad3m
         (205 letters)

gb|AA278224|AA2278224 zs77e05.r1 NCI_CGAP_GCB1 Homo sapiens cDNA clone IMAGE:703520 5'
          similar to TR:G1184157 G1184157 MAX-INTERACTING
          TRANSCRIPTIONAL REPRESSOR. ;
          Length = 430

 Score =  209  bits (526), Expect = 1e-53
 Identities =  104/124 (83%), Positives = 116/124 (92%), Gaps = 1/124 (0%)
 Frame = +2

Query: 1    MEPVASNIQVLLQAAEFLERREREAEHGYASLCPHHSPGTVCRRRKPPLQAPGALNSGRS  60
             MEP+ASNIQVLLQAAEFLERREREAEHGYASLCPH SPG + RR+K P QAPGA +SGRS
Sbjct: 56   MEPLASNIQVLLQAAEFLERREREAEHGYASLCPHRSPGPIHRRKKRPPQAPGAQDSGRS  235

Query: 61   VHNELEKRRRAQLKRCLEQLRQQMPLGVDCTRYTTLSLL-RARVHIQKLEEQEQQARRLK  119
             VHNELEKRRRAQLKRCLE+L+QQMPLG DC RYTTLSLL RAR+HIQKLE+QEQ+AR+LK
Sbjct: 236  VHNELEKRRRAQLKRCLERLKQQMPLGGDCARYTTLSLLRRARMHIQKLEDQEQRARQLK  415

Query: 120  EKLRS 124
             E+LR+
Sbjct: 416  ERLRT 430

dbj|C02407|C02407 HUMGS0012279, Human Gene Signature, 3'-directed cDNA sequence.
          Length = 348

 Score = 97.5 bits (239), Expect = 6e-20
 Identities = 51/63 (80%), Positives = 56/63 (87%)
 Frame = +3

Query: 125 KQQSLQQQLEQLQGLPGARERERLRADSLDSSGLSSERSDSDQEDLEVDVENLVFGTETE  184
            KQQSLQ+   QL+GL GA ERERLRADSLDSSGLSSERSDSDQE+LEVDVE+LVFG E E
Sbjct: 45  KQQSLQRXWMQLRGLAGAAERERLRADSLDSSGLSSERSDSDQEELEVDVESLVFGGEAE  224

Query: 185 LLQ 187
            LL+
Sbjct: 225 LLR 233
```

FIG.17A

BASE COUNT 130 a 234 c 258 g 106 t 5 others
ORIGIN

```
  1  cagccgcttg ctccggccgg caccctaggc cgcagtccgc caggctgtcg ccgacatgga
 61  acccttggcc agcaacatcc aggtcctgct gcaggcggcc gagttcctgg agcgccgtga
121  gagagaggcc gagcatggtt atgcgtccct gtgcccgcat cgcagtccag gccccatcca
181  caggaggaag aagcgacccc cccaggctcc tggcgcgcag gacagcgggc ggtcagtgca
241  caatgaactg gagaagcgca ggagggccca gttgaagcgg tgcctggagc ggctgaagca
301  gcagatgccc ctgggcggcg actgtgcccg gtacaccacg ctgagcctgc tgcgccgtgc
361  caggatgcac atccagaagc tggaggatca ggagcagcgg gcccgacagc tcaaggagag
421  gctgcgcaca aagcagcaga gcctgcagcg gcantggatg cagctccggg ggctggcagg
481  ngcggccgag cgggagcgnc tgcgggcgga cagtctggac tcctcaggcc tctcctctga
541  gcgctcagac tcagaccaag aggagctgga ggtggatgtg gagagcctgg tgtttggggg
601  tgaggccgag ctgctgcggg gcttcgtcgc cggccaggag cacagctact cgcacgtcgg
661  cggcgcctgg ctatgatgtt cctcacccan ggcgggcctc tgccctctta ctcgttgccc
721  aagcccactt tnc
```

FIG.17B

>*Mad*3b(Putative)
MEPLASNIQVLLQAAEFLERREREAEHGYASLCPHRSPGPIHRRKKRPPQAPGAQDSGRSVHNELEKRRRAQLK
RCLERLKQQMPLGGDCARYTTLSLLRRARMHIQKLEDQEQRARQLKERLRTKQQSLQRXWMQLRGLAGAAERER
LRADSLDSSGLSSERSDSDQEELEVDVESLVFGGEAELLRGFVAGQEHSYSHVGGAWL

FIG.17C

```
gi|2506888|MADe      MATAVGHNIQLLLEAADYLERREREAEHGYASMLPYS-KDADAFKRRNKPKKNST--SSRSTHNEMEKNRRAHLRLCLEKLKGLVPLGPESSRHTTLSLL
gi|729978|MADh       MAAAYRMNIQMLLEAADYLERREREAEHGYASMLPYNNKDRDALKRRNKSKKNNS--SSRSTHNEMEKNRRAHLRLCLEKLLGLVPLGPESSRHTTLSLL
gi|2792362|Mad4h     ---MELNSLLILLEAAEYLERRDREAEHGYASVLPFDGDFAREKTKAAGLVRKAP--NNRSSHNELEKHRRAKLRLYLEQLKQLVPLGPDSTRHTTLSLL
gi|217199|Mad4m      ---MELNSLLLLLEAAEYLERRDREAEHGYASMLPFDGDFARKKTKTAGLVRKGP--NNRSSHNELEKHRRAKLRLYLEQLKQLGPLGPDSTRHTTLSLL
gi|2137199|Mad3m     -MEPVASNIQVLLQAAEFLERREREAEHGYASLCPHHSPGTVCRRRKPPLQAPGALNSGRSVHNELEKRRRAQLKRCLEQLRQQMPLGVDCTRYTTLSLL
Mad3h Putative       -MEPLASNIQVLLQAAEFLFRREREAEHGYASLCPHRSPGPIHRRKKRPPQAPGAQDSGRSVHNELEKRRRAQLKRCLERLKQQMPLGGDCARYTTLSLL

gi|2506888|MADe      TKAKLHIKKLEDCDRKAVHQIDQLQREQRHLKRRLEKLGAERTR-------MDSVG-SVVSSERSDSDREELDVDVDVDVDVDVEGTDYLPGDLGWSSS-
gi|729978|MADh       TKAKLHIKKLEDCDRKAVHQIDQLQREQRHLKRQLEKLGIERIR-------MDSIG-STVSSERSDSDRE--------EIDVDVESTDYLTGDLDWSSSS
gi|2792362|Mad4h     KRAKVHIKKLEEQQRRALSIKEQLQQEHRFLKRRLEQLSVQSVER----VRTDSTG-SAVSTD--DSEQE----------VDIEGMEFGPGELDSVGS-
gi|2137499|Mad4m     K-AKMHIKKLEEQQRRALSIKEQLQREHRFLKRRLEQLSVQSVR-----VRTDSTG-SAVSTD--DSEQE----------VDIEGMEFGPGELDSAGS-
gi|2137498|Mad3m     R-ARVHIQKLEEQEQQARRLKEKLRSKQQSLQQQLEQLQGLPGARERPRLRADSLDSSGLSSERSDSDQE--------DLSVDVENLVFG-TETELLQSF
Mad3h Putative       RRARMHIQKLEDQEQRARQLKERLRTKQQSLQRXWMQLRGLAGAAERERLRADSLDSSGLSSERSDSDQE--------ELEVDVESLVFG-GEAELLRGF

gi|2506888|MADe      VSDSDERGSMQSLG-SDEGYSSATVKRAKLQQGHKAGLGL
gi|729978|MADh       VSDSDERGSMQSLG-SDEGYSSTSIKRIKLQQSHKACLGL
gi|2792362|Mad4h     SSDADDHYSLQSGTGGDSGFGPHCRRLGRPALS-------
gi|2137499|Mad4m     SSDADDHYSLQSSGCSDSSYGHPCRRPGCPGLS-------
gi|2137498|Mad3m     SAGREHSYSHSTCAWL----------------------
Mad3h Putative       VAGQEHSYSHVGGAWL----------------------
```

FIG.17D

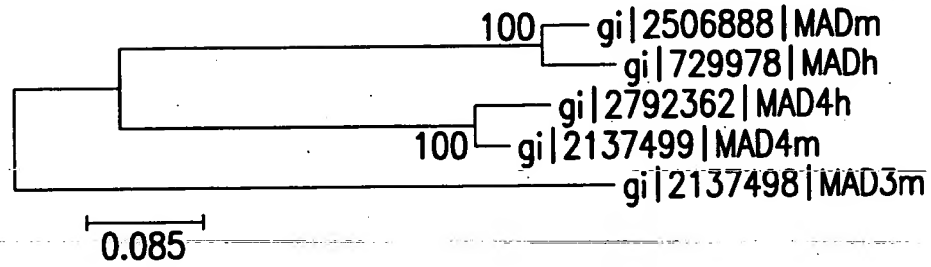


FIG.18A

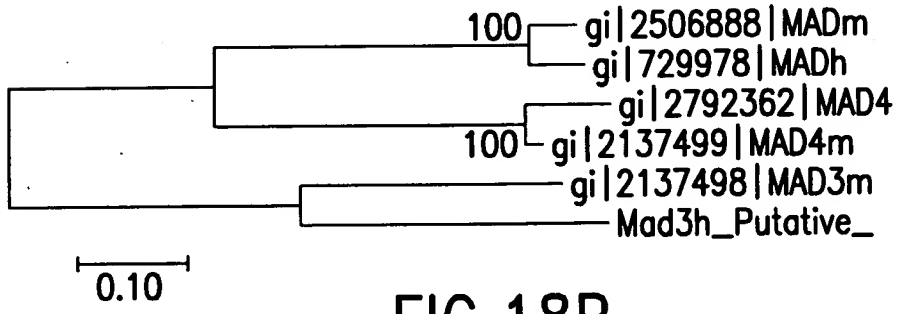


FIG.18B